

Energy Costs for Traffic Offloading by Cache-enabled D2D Communications

Binqiang Chen and Chenyang Yang
Beihang University, Beijing China
Email: chenbq@buaa.edu.cn, cyyang@buaa.edu.cn

Abstract—Device-to-Device (D2D) communications can offload the traffic and boost the throughput of cellular networks. By caching files at users, content delivery traffic can be offloaded via D2D links, if a helper user are willing to send the cached file to the user who requests the file. Yet it is unclear how much energy needs to be consumed at a helper user to support the traffic offloading. In this paper, we strive to find the minimal energy consumption required at a helper user to maximize the amount of offloaded traffic. To this end, we introduce a user-centric proactive caching policy that can control the energy cost for a helper user to convey a file, and then optimize the caching policy to maximize the offloaded traffic. To reduce the energy during transmission, we optimize the transmit power to minimize the energy consumed by a helper to send a file. We analyze the relationship between traffic offloading and energy cost with the optimized caching policy and transmit power by numerical and simulation results, which demonstrate that a significant amount of traffic can be offloaded with affordable energy costs.

Index Terms—Caching, D2D, Traffic offloading, Energy cost.

I. INTRODUCTION

Device-to-device (D2D) communications is a promising approach to offload the traffic and boost the throughput of cellular networks, whose typical user cases include content distribution, gaming and relaying [1].

The lion's share of cellular traffic is video distribution, which will generate more than 69% of mobile data traffic by 2019 [2]. Nonetheless, a large amount of traffic is generated by a few contents. On the other hand, the storage of mobile devices grows rapidly with low cost. Motivated by these facts, cache-enabled D2D communications was proposed in [3, 4] to offload the traffic of video transmission. Without caching at the devices, the users need to fetch their desired content via base station (BS) from a remote server. By pre-caching popular files at users during the off-peak time according to their possible interests, the desired file of a user can be transmitted via D2D links by the users in proximity when the actual request arrives. Such a proactive caching policy largely alleviates the burden at the BSs during the peak time. To maximize the traffic offloaded by D2D links, the policy to proactively cache popular files at mobile devices was optimized in [5], and a distributed reactive caching mechanism was designed in [6].

Different from the D2D use cases of supporting peer-to-peer services such as gaming, where the users acting as transmitters

are willing to send messages to the destination users [7], offloading the content delivery traffic by cache-enabled D2D communications needs the help of other users who have cached the desired files but are not obligated to act as helpers to transmit the files. Due to the limited battery capacity, a natural question from a helper user in such a network is: “why should I spend energy of my battery to provide you with faster video download? [3]” This makes the energy consumption at a helper user a big concern in D2D communications with caching. Considering the large potential in traffic offloading by cache-enabled D2D communications, it is an urgent task to evaluate the energy consumed at a helper user to deliver the files. In particular, characterizing the required energy consumption at a helper to maximize the traffic offloading for the network is an important problem.

In previous research efforts [3–6], the energy costs at helper users are never considered. On one hand, maximal transmit power is always used to deliver the files. On the other hand, by dividing the users in a cell into clusters and assuming that only the users within a cluster can establish D2D links, the optimal caching policy proposed in [5] cannot maximize the offloaded traffic in a cell, and the energy cost for a helper is high. This is because when a user is only allowed to communicate with the users within its cluster, it can not fetch the file from the nearest helper in adjacent cluster who can convey the file with low transmit power.

In this paper, we strive to investigate the energy cost at a helper user spent to maximize the offloaded traffic. Toward this goal, we first introduce a user-centric proactive caching policy, where only the users within a *collaboration distance* of a user can serve as helpers. We optimize the caching policy to maximize traffic offloading with a given collaboration distance and the user demands statistics. When the collaboration distance is large, the probability that the users can fetch their desired contents via D2D links is high, and then more traffic can be offloaded. However, since the possible D2D link distance increases, the energy cost of a helper user also grows. By setting the collaboration distance according to the affordable energy cost of each helper, such a caching policy helps reduce the energy for transmission. Then, we optimize the transmit power for a helper to send the requested file to minimize the energy spent by the helper. Finally, we analyze the tradeoff between traffic offloading and energy costs for the D2D network with optimized caching policy and transmit

power with numerical and simulation results.

II. SYSTEM MODEL

Consider one cell in cellular networks where user location follows a Poisson Point Process (PPP) with density λ . Each single antenna user has local cache to store files and acts as helper. If a helper establishes a D2D link with a D2D receiver (DR), it becomes a D2D transmitter (DT). For notational simplicity, we assume that each user stores one file as in [4]. The BS is aware of the cached files of the users and coordinates the D2D communications.

We consider a static content catalog including N_f files that all users in the cell may request, where the files are indexed according to the popularity, and the 1st file is the most popular file. Each file is with size of F Bits. Each user requests a file from the catalog independently. The probability that the i th file is requested by a user is assumed to follow Zipf distribution, which is

$$p_r(i) = i^{-\beta} / \sum_{k=1}^{N_f} k^{-\beta}, \quad (1)$$

where $\sum_{i=1}^{N_f} p_r(i) = 1$, and the parameter β reflects the popularity of the files [8].

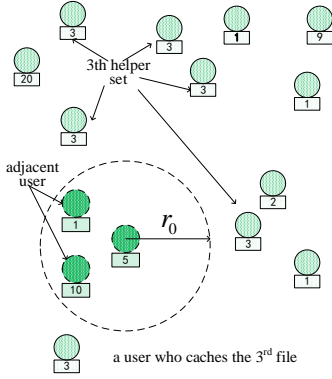


Fig. 1. Illustration for a cache-enabled D2D network

Since transmitting to a far away DR spends more energy of a DT, we introduce a user-centric caching policy to control the energy for transmission: a helper user will send the file it cached to the user requesting the file only if their distance is smaller than a given value r_c , called *collaboration distance*. We consider a probabilistic caching policy, where each user caches a file according to a r_c -dependent caching distribution, i.e., the probability that the i th file is cached at users, $i = 1, \dots, N_f$. All users in the cell cached with the i th file constitute a user set, called the i th *helper set*. In practice, the files can be proactively downloaded by the BS during the off-peak time.

The users with distance r less than r_c are called *adjacent users*, as shown in Fig. 1. If a user can find its requested file in the local caches of its adjacent users, a D2D link will be established between the user and its nearest adjacent user cached with the requested file to convey the file. Otherwise, the user needs to fetch the file from the BS. The probability that the requests from the users can be served via D2D links is

called *traffic offloading ratio*, which reflects how much traffic can be offloaded by D2D communications.

III. OPTIMAL CACHING POLICY

In this section, we optimize the caching distribution to maximize the offloading ratio.

Denote the probability that the i th file is cached at users as $p_c(i)$. Then, the locations of the users who belong to the i th *helper set* follow a PPP distribution with density $\lambda p_c(i)$ [9]. Thus, the probability that a user requesting the i th file can find its desired file in the cache of any user within the collaboration distance r_c is

$$p_f(i) = 1 - e^{-\lambda p_c(i) \pi r_c^2}. \quad (2)$$

Then, the offloading ratio with given caching distribution and collaboration distance can be obtained from (1) and (2) as

$$p_o(p_c(i), r_c) = \sum_{i=1}^{N_f} p_r(i) (1 - e^{-\lambda p_c(i) \pi r_c^2}). \quad (3)$$

The optimal caching distribution that maximizes the offloading ratio can be found from the following problem

$$\begin{aligned} \max_{p_c(i)} \quad & p_o(p_c(i), r_c) \\ \text{s.t.} \quad & \sum_{i=1}^{N_f} p_c(i) = 1, \quad p_c(i) \geq 0, \quad i = 1, \dots, N_f. \end{aligned} \quad (4)$$

Because the objective function is the sum of N_f exponential functions with linear constraints, this problem is convex [10]. According to the Karush-Kuhn-Tucker (KKT) conditions of this problem, the optimal caching distribution should satisfy the following conditions

$$\begin{aligned} p_c^*(i) &= \left[\frac{1}{\lambda \pi r_c^2} \ln(p_r(i)) - \frac{1}{\lambda \pi r_c^2} \ln\left(\frac{-\mu}{\pi \lambda r_c^2}\right) \right]^+, \quad \forall i, \\ \sum_{i=1}^{N_f} p_c^*(i) &= 1, \end{aligned} \quad (5)$$

where $[x]^+ = \max(x, 0)$.

Proposition 1: If $\frac{(N_f)^{N_f}}{N_f!} < e^{\frac{\lambda \pi r_c^2}{\beta}}$, then the optimal caching distribution will be

$$p_c^*(i) = \frac{\beta}{\lambda \pi r_c^2 N_f} \sum_{j=1}^{N_f} \ln\left(\frac{j}{i}\right) + \frac{1}{N_f}. \quad (6)$$

Otherwise, the optimal caching distribution will be

$$p_c^*(i) = \begin{cases} \frac{\beta}{\lambda \pi r_c^2} \sum_{j=1}^{i^*} \ln\left(\frac{j}{i}\right) + \frac{1}{i^*}, & i \leq i^* \\ 0, & i^* < i \leq N_f \end{cases}, \quad (7)$$

where i^* satisfies $\frac{(i^*+1)^{i^*}}{i^*!} \geq e^{\frac{\lambda \pi r_c^2}{\beta}}$, $\frac{(i^*)^{i^*}}{i^*!} < e^{\frac{\lambda \pi r_c^2}{\beta}}$ and $\frac{\lambda \pi r_c^2}{\beta} - 1 \leq i^* \leq \frac{\lambda \pi r_c^2}{\beta} + \ln(\sqrt{2\pi N_f}) + 1$.

Proof: See Appendix A ■

With Prop. 1, the optimal caching distribution $p_c^*(i)$, $i = 1, \dots, N_f$ can be obtained efficiently, which depends on the collaboration distance r_c , user density λ , as well as content

statistics N_f and β . With the optimized caching distribution, each user can randomly select a file to cache according to the probability $p_c^*(i)$. Then, the traffic in the network can be maximally offloaded. Because such a caching policy depends on r_c , the possible D2D link distance and hence the energy spent by a helper can be controlled by the pre-determined value of collaboration distance.

IV. ENERGY CONSUMPTION FOR TRAFFIC OFFLOADING

In this section, we investigate the energy consumed by each DT for offloading the traffic by the cache-enabled D2D communications. To this end, we first optimize the transmit power for conveying a file to minimize the energy consumption at each DT. Then, we evaluate the average energy consumed at each DT to transmit a file via D2D links based on the optimal transmit power and optimal caching policy.

Once a D2D link is established, the DT can transmit its cached file to the DR who requests the file. Considering the random user requests, random caching and random user locations, it is reasonable to treat the interference among the D2D links as noise. Then, the signal to interference plus noise ratio (SINR) at the DR can be expressed as

$$\gamma(r) = \frac{P_t |h|^2 r^{-\alpha}}{\sigma_0^2}, \quad (8)$$

where P_t is the transmit power at the DT, σ_0^2 is the variance of white Gaussian noise and interference, h and r are respectively the channel coefficient and distance between the DT and the DR with h following zero mean complex Gaussian distribution with unit variance, and α is the path loss exponent.

Considering that $1 + \gamma(r)$ approximately follows Gamma distribution [11], the average data rate with respect to small scale channel fading can be derived as

$$\bar{R}(r) = \mathbb{E}\{W \log_2(1 + \frac{P_t |h|^2 r^{-\alpha}}{\sigma_0^2})\} \approx W \log_2(1 + \frac{P_t r^{-\alpha}}{\sigma_0^2}), \quad (9)$$

where W is the bandwidth.

By using the first order approximation [12], the average time to convey a file of size F can be approximated as $F/\bar{R}(r)$. Then, the energy consumed to transmit a file via a D2D link with distance r can be approximated as

$$E_c(r) \approx \frac{F}{W \log_2(1 + \frac{P_t(r) r^{-\alpha}}{\sigma_0^2})} \left(\frac{1}{\eta} P_t(r) + P_c \right), \quad (10)$$

where P_c is the circuit power consumed at the DT, and η is the power amplifier efficiency.

To minimize the energy consumption, the optimal transmit power at the DT can be obtained from the following problem,

$$\begin{aligned} \min_{P_t(r)} \quad & \frac{F}{W \log_2(1 + \frac{P_t(r) r^{-\alpha}}{\sigma_0^2})} \left(\frac{1}{\eta} P_t(r) + P_c \right) \\ \text{s.t.} \quad & 0 < P_t(r) \leq P_{\max}, \end{aligned} \quad (11)$$

where P_{\max} is the maximal transmit power of the DT.

Proposition 2: Denote $y = 1 + P_t(r) \frac{r^{-\alpha}}{\sigma_0^2}$, $y_0 = 1 + P_{\max} \frac{r^{-\alpha}}{\sigma_0^2}$, and $\epsilon = \frac{r^{-\alpha} \eta P_c}{\sigma_0^2} - 1$. If $(\frac{y_0}{e})^{y_0} < 2^{\epsilon/\ln 2}$, then

the optimal solution of problem (11) will be $P_t^*(r) = P_{\max}$. Otherwise, $P_t^*(r)$ will satisfy the following condition

$$\left(\frac{y}{e}\right)^y = 2^{\epsilon/\ln 2}. \quad (12)$$

Proof: See Appendix B. ■

With Prop. 2, the optimal transmit power $P_t^*(r)$ can be obtained efficiently. By substituting $P_t^*(r)$ to (10), the minimal energy consumption for the D2D link with distance r can be obtained as $E_c^*(r)$.

To evaluate minimal energy cost for a DT to transmit a file over D2D links with different distances, we need to obtain the distribution of r when the optimal caching policy is employed. Similar to (2), the cumulative distribution function of the distance between the DR requesting the i th file and its nearest DT in the i th helper set with optimized caching policy can be obtained as $F(p_c^*(i), r) = 1 - e^{-\lambda p_c^*(i) \pi r^2}$. Therefore, the probability density function of the D2D link distance can be obtained as

$$f(p_c^*(i), r) = \frac{dF(p_c^*(i), r)}{dr} = 2\pi r \lambda p_c^*(i) e^{-\lambda p_c^*(i) \pi r^2}. \quad (13)$$

Then, for a given collaboration distance r_c , the average energy consumed at the DT with the optimized transmit power can be obtained as

$$\begin{aligned} \bar{E}_c^*(r_c) &= \sum_{i=1}^{N_f} p_r(i) \int_0^{r_c} E_c^*(r) f(p_c^*(i), r) dr \\ &= \int_0^{r_c} E_c^*(r) p'_o(p_c^*(i), r_c) dr \end{aligned} \quad (14)$$

where $p'_o(p_c^*(i), r_c)$ is the first-order derivative of $p_o(p_c^*(i), r_c)$ with respect to r_c , and $p_o(p_c^*(i), r_c)$ is obtained from (3) with the optimized caching distribution.

Proposition 3: Both the maximal offloading ratio $p_o(p_c^*(i), r_c)$ and the minimal average energy cost $\bar{E}_c^*(r_c)$ increase with the collaboration distance r_c .

Proof: See Appendix C. ■

Prop. 3 implies that there is a tradeoff between the traffic offloading and the energy consumption.

V. SIMULATION AND NUMERICAL RESULTS

In this section, we evaluate the accuracy of the approximations and the energy consumption at a DT for offloading via simulation and numerical results.

To show the impact of the optimized caching policy and transmit power, we consider uniform caching policy (i.e., all users select a file from the catalog uniformly) as a *caching baseline*, and a transmit policy always using the maximal transmit power as the *transmission baseline*.

We consider a square cell with side length 1000m. The users' location follows PPP distribution with $\lambda = 0.03$, then there are 2 ~ 3 users in a 10m × 10m area. The path-loss model is $37.6 + 36.8 \log_{10}(r)$, where r is the distance of D2D link [5]. $W = 20$ MHz and $\sigma_0^2 = -95$ dBm. The file catalog is with $N_f = 1000$ files, where each file is with size of 30 Mbytes (i.e., a typical video size on the Youtube website [3]). The

parameter of Zipf distribution $\beta = 1$. The maximal transmit power at DT is $P_{\max} = 23$ dBm, the amplifier efficiency $\eta = 0.2$, and the circuit power is $P_c = 115.9$ mW [13].

To reflect how much average energy consumed at a DT to transmit a file occupies the battery capacity, we evaluate an *energy cost ratio* as follows,

$$\bar{p}_E = \frac{\bar{E}_c(r_c)}{3.6V_0Q}, \quad (15)$$

where Q is the battery capacity in mAh, and V_0 is the operating voltage in V.

The operating voltage at user is set as $V_0 = 4V$ and the battery capacity $Q = 1800$ mAh (typical for iPhone 6).

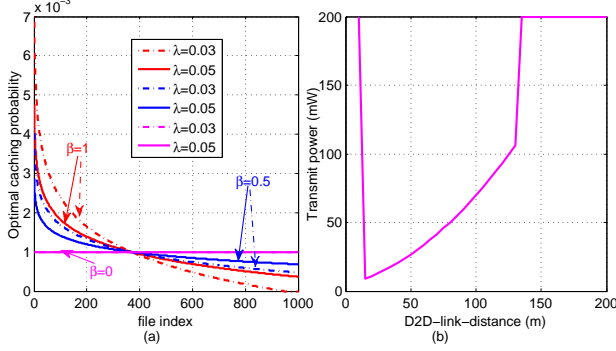


Fig. 2. Optimal caching distribution and optimal transmit power

In Fig. 2, we show the optimal caching distribution and the optimal transmit power versus the D2D link distance r . We can see from Fig. 2(a) that the optimal caching distribution is similar to the Zipf distribution, where the file with smaller index (i.e., more popular) has higher probability to be cached, which agrees with the intuition. With the increase of β and λ , the caching probability for popular files increases. We can see from Fig. 2(b) that the optimal transmit power for very small and large D2D link distance r is the maximal transmit power. This is because when r is small, the transmit duration can be reduced and the energy cost can be minimized with P_{\max} . When r is large, the circuit power and transmit power are balanced to minimize the energy cost with P_{\max} .

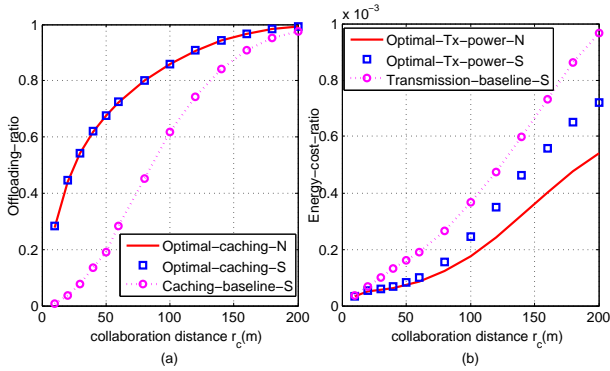


Fig. 3. Offloading ratio and average energy costs versus r_c , $\beta = 1$, $\lambda = 0.03$, S-Simulation results, N-Numerical results

In Fig. 3, we show the offloading ratio and the energy cost ratio versus the collaboration distance r_c , where the energy

consumption at each DT is computed for the cache-enabled D2D communications with optimized caching distribution. The simulation results are close to numerical results when $r_c < 100$ m, which indicates that the approximated energy consumption is accurate for smaller collaboration distance. As expected, the optimized caching policy can offload more traffic than the baseline. Even when $r_c = 10$ m, more than 20% traffic can be offloaded with the optimal caching policy. When r_c is large, e.g., $r_c = 200$ m, the offloading ratios of both the optimal and baseline caching policy approach to one, since adjacent users for each user is abundant to cache almost all the files in the catalog. The optimized transmit power consumes less energy than always using the maximal transmit power, especially when the collaboration distance is large. From this figure, we can pre-determine the collaboration distance to control the energy cost at each helper into an affordable level.

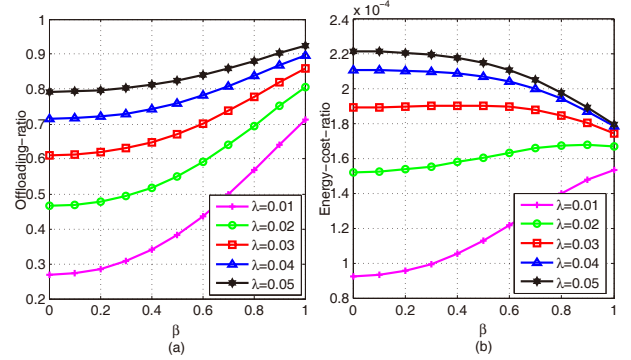


Fig. 4. Offloading ratio and average energy costs versus β and λ , $r_c = 100$ m

In Fig. 4, we show the offloading ratio with optimal caching policy and the energy cost ratio with optimal transmit power versus Zipf distribution parameter β and user density λ . With the increase of λ , a user has more adjacent users, therefore, the probability that a requested file can be transmitted via D2D link increases. As a result, the energy cost ratio increases. The offloading ratio increases with β , and grows rapidly for heavy traffic load. However, with the growth of β , the energy cost ratio increases when λ is small but decreases when λ is large.

In Fig. 5(a), we show the tradeoff between the offloading ratio and the energy cost ratio with different caching and power control policies, where the offloading ratio is adjusted by changing the collaboration distance r_c from 10m to 100m. We can see that with the optimized caching policy and transmit power, the energy cost ratio increases with offloading ratio slowly, and is much lower than other policies. To offload 80% traffic, the energy consumption at each DT only occupies 0.02% battery capacity. This suggests that D2D communications with caching will be cost-efficient for offloading if the caching and transmission policies are judiciously designed.

In the sequel, we show the impact of several key factors on the tradeoff between the offloading ratio and the energy cost ratio with the optimized caching policy and transmit power.

In Fig. 5(b), we show the impact of interference level. As expected, with the increase of interference level, to achieve the same offloading ratio, more energy needs to be consumed at

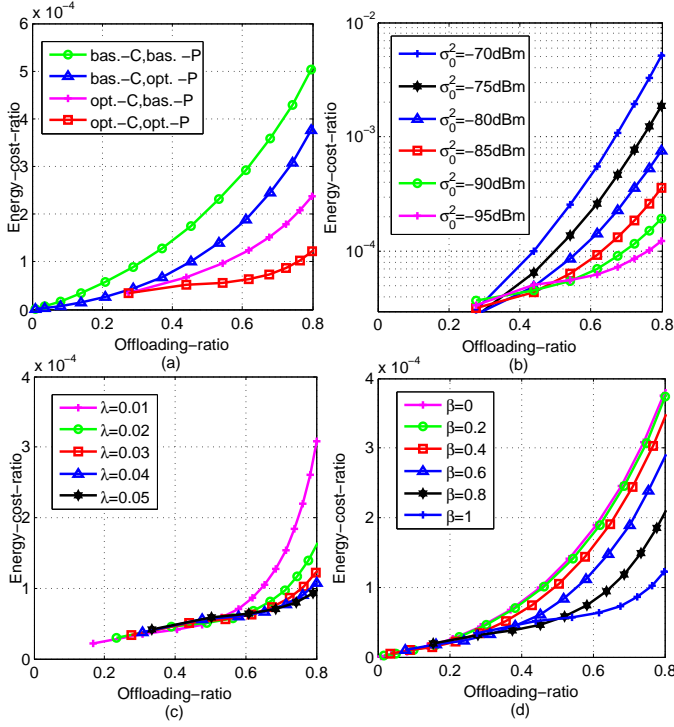


Fig. 5. Tradeoff between offloading ratio and energy cost ratio, “opt.”: optimal, “bas.”: baseline, “C”: caching, “P”: power control, where $\lambda = 0.03$, $\beta = 1$, $\sigma_0^2 = -95\text{dBm}$ if not specified.

each DT. When $\sigma_0^2 = -70\text{ dBm}$, to achieve an 80% offloading ratio, the energy cost ratio is 1%. This suggests the importance of controlling the interference among D2D links.

In Fig. 5(c), we show the impact of the users density λ . As λ increases, the number of adjacent users of each user increases, hence less energy cost is consumed to achieve the same offloading ratio, resulting in the reduction of the energy cost ratio. Moreover, the increasing speed of energy cost with the offloading ratio becomes slowly when λ is large. This suggests that it is more efficient in terms of the energy cost of each DT for offloading by cache-enabled D2D communications in the network with heavy traffic load.

In Fig. 5(d), we show the impact of Zipf distribution β . As expected, with the increase of β , less energy is consumed to achieve a given offloading ratio. To achieve an 80% traffic offloading ratio, the energy cost of each DT for the scenario with $\beta = 1$ is only 25% of that with $\beta = 0$.

VI. CONCLUSION

In this paper, we strive to evaluate the energy consumed at a helper user to support traffic offloading for a network by cache-enabled D2D communications. We first optimized a user-centric proactive caching policy, with which the traffic can be maximally offloaded and the energy consumed for transmission can be controlled by a collaboration distance. We then optimized the transmit power to convey a file via D2D link, aimed to minimize the energy consumption at the helper user for any given caching policy. With the optimal caching distribution and optimal transmit power, we investi-

gated the tradeoff between traffic offloading and energy cost. Simulation and numerical results showed that the traffic can be significantly offloaded by the cache-enabled D2D links with little energy costs at each help user.

APPENDIX A: PROOF OF PROPOSITION 1

Denote

$$x_i \triangleq \frac{\ln(p_r(i))}{\lambda \pi r_c^2}, \text{ and } v \triangleq \frac{1}{\lambda \pi r_c^2} \ln\left(\frac{-\mu}{\pi \lambda r_c^2}\right). \quad (16)$$

Then, by substituting the first condition into the second condition, the necessary condition in (5) can be rewritten as

$$\sum_{i=1}^{N_f} [x_i - v]^+ = 1. \quad (17)$$

Since problem (4) is convex, the solution of v found from this condition is global optimal, with which the optimal caching distribution can be obtained from (5).

Because $p_r(i)$ is an decreasing function of file index i from (1), and $p_c^*(i)$ decreases with $p_r(i)$ from (5), $p_c^*(i)$ is an decreasing function of i . Thus, there exists a unique file index $i^* \leq N_f$, where $p_c^*(i) > 0$ when $i \leq i^*$, $p_c^*(i) = 0$ otherwise. Finding the solution of v from (17) is equivalent to finding the index i^* from

$$\sum_{i=1}^{i^*} (x_i - v) = 1, \quad (18)$$

since once i^* is found, v^* can be obtained as

$$v^* = \frac{\sum_{i=1}^{i^*} x_i - 1}{i^*}. \quad (19)$$

- Case 1: When $i^* = N_f$, $p_c^*(N_f) = x_{N_f} - v > 0$, which can be rewritten as $\sum_{i=1}^{N_f} (x_i - x_{N_f}) < 1$ after substituting v in (19). Considering (16) and (1), we have

$$\begin{aligned} \sum_{i=1}^{N_f} (x_i - x_{N_f}) &= \sum_{i=1}^{N_f} \frac{\ln(p_r(i)) - \ln(p_r(N_f))}{\lambda \pi r_c^2} \\ &= \frac{\beta}{\lambda \pi r_c^2} \sum_{i=1}^{N_f} \ln(N_f/i) = \frac{\beta}{\lambda \pi r_c^2} \ln\left(\frac{N_f^{N_f}}{N_f!}\right) < 1, \end{aligned} \quad (20)$$

which can be rewritten as $\frac{(N_f)^{N_f}}{N_f!} < e^{\frac{\lambda \pi r_c^2}{\beta}}$. When this inequality holds, $i^* = N_f$.

By substituting v^* in (19) into (5), the optimal caching distribution can be derived as

$$p_c^*(i) = \frac{\beta}{\lambda \pi r_c^2 N_f} \sum_{j=1}^{N_f} \ln\left(\frac{j}{i}\right) + \frac{1}{N_f}. \quad (21)$$

- Case 2: When $i^* < N_f$, $p_c^*(i^*) = x_{i^*} - v > 0$ and $x_{i^*+1} - v \leq 0$. By substituting v in (19) into these two inequalities, we have

$$\sum_{i=1}^{i^*} (x_i - x_{i^*+1}) \geq 1, \quad \sum_{i=1}^{i^*} (x_i - x_{i^*}) < 1, \quad (22)$$

which can be further derived by considering $p_r(i)$ in (1) and x_i in (16) as

$$\frac{\beta}{\lambda\pi r_c^2} \ln\left(\frac{(i^*+1)^{i^*}}{i^*!}\right) \geq 1, \quad \frac{\beta}{\lambda\pi r_c^2} \ln\left(\frac{(i^*)^{i^*}}{i^*!}\right) < 1. \quad (23)$$

Then, i^* satisfies $\frac{(i^*+1)^{i^*}}{i^*!} \geq e^{\frac{\lambda\pi r_c^2}{\beta}}$ and $\frac{(i^*)^{i^*}}{i^*!} < e^{\frac{\lambda\pi r_c^2}{\beta}}$. By substituting v in (19) into (5), we obtain

$$p_c^*(i) = \begin{cases} \frac{\beta}{\lambda\pi r_c^2 i^*} \sum_{j=1}^{i^*} \ln\left(\frac{j}{i}\right) + \frac{1}{i^*}, & i \leq i^* \\ 0, & i^* < i \leq N_f. \end{cases} \quad (24)$$

With Stirling formula [14], $\sqrt{2\pi n} \left(\frac{n}{e}\right)^n < n! < \sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\frac{1}{12}}$, (23) can be further derived as

$$i^* - \ln(\sqrt{2\pi i^*} e^{\frac{1}{12}}) < \frac{\lambda\pi r_c^2}{\beta}, \quad i^* + 1 - \ln(\sqrt{2\pi i^*}) > \frac{\lambda\pi r_c^2}{\beta}.$$

Since $i^* < N_f$, $\ln(\sqrt{2\pi i^*}) < \ln(\sqrt{2\pi N_f})$, we can obtain the range of i^* as $\frac{\lambda\pi r_c^2}{\beta} - 1 \leq i^* \leq \frac{\lambda\pi r_c^2}{\beta} + \ln(\sqrt{2\pi N_f}) + 1$.

APPENDIX B: PROOF OF PROPOSITION 2

By denoting $z = P_t(r)$, $a = \frac{r^{-\alpha}}{\sigma_0^2}$, $b = \eta P_c$ and $A = \frac{F}{W\eta}$, the objective function in (11) can be rewritten as

$$f(z) = A \frac{z + b}{\log_2(1 + az)}, \quad (25)$$

where $z, a, b > 0$ and $z \leq P_{max}$. By taking the first-order derivative of $f(z)$, we obtain

$$f'(z) = A \frac{((1 + az) \log_2(1 + az) \ln 2 - (1 + az) - \epsilon)}{(1 + az) \log_2^2(1 + az) \ln 2}. \quad (26)$$

Denote $g(y) = y \log_2(y) \ln 2 - y - \epsilon$, where $y = 1 + az$, $1 < y \leq 1 + aP_{max} \triangleq y_0$, and $\epsilon = \frac{r^{-\alpha} \eta P_c}{\sigma_0^2} - 1$. We can see that if $g(y) > 0$, then $f'(z) > 0$, and vice versa.

By taking the first-order derivative of $g(y)$, we obtain $g'(y) = \log_2(y) \ln 2 > 0$ due to $y > 1$. Therefore, $g(y)$ is an increasing function of y , when $y \rightarrow 1$, $g(y) \rightarrow 0$. If $g(y_0) \leq 0$, $g(y)$ is always less than zero since $1 < y < y_0$, then $f(z)$ will be a decreasing function, i.e., $f(z)$ achieves its minimum when $z = P_{max}$. Otherwise, $f(z)$ first decreases and then increases, and achieves its minimum when $g(y) = 0$.

In summary, if $g(1 + \frac{P_{max} r^{-\alpha}}{\sigma_0^2}) \leq 0$, $P_t^*(r) = P_{max}$. Otherwise, we can find $P_t^*(r)$ efficiently by searching from $(\frac{y}{e})^y = 2^{\epsilon/\ln 2}$, which is obtained from $g(y) = 0$.

APPENDIX C: PROOF OF PROPOSITION 3

Consider two collaboration distances $r'_c > r_c$, and the corresponding optimal caching distribution are respectively $p_c^*(i)'$ and $p_c^*(i)$. We can see from (3) that $p_o(p_c^*(i), r)$ is an increasing function of r ($0 < r \leq r_c$). Then, $p_o(p_c^*(i)', r'_c) \geq p_o(p_c^*(i), r'_c) \geq p_o(p_c^*(i), r_c)$. Thus, $p_o(p_c^*(i), r_c)$ is an increasing function of r_c .

By subtracting $\bar{E}_c^*(r_c)$ from $\bar{E}_c^*(r'_c)$, we have

$$\begin{aligned} & \bar{E}_c^*(r'_c) - \bar{E}_c^*(r_c) \\ &= \int_0^{r'_c} E_c^*(r) p'_o(p_c^*(i)', r) dr - \int_0^{r_c} E_c^*(r) p'_o(p_c^*(i), r) dr \\ &\geq \int_0^\xi E_c^*(r) \Delta p_o dr + \int_\xi^{r'_c} E_c^*(r) \Delta p_o dr \\ &\geq E_c^*(\xi) \left(\int_\xi^{r'_c} \Delta p_o dr + \int_0^\xi \Delta p_o dr \right) \\ &= E_c^*(\xi) \left(\int_0^{r'_c} p'_o(p_c^*(i)', r) dr - \int_0^{r'_c} p'_o(p_c^*(i), r) dr \right) \\ &= E_c^*(\xi) (p_o(p_c^*(i)', r) - p_o(p_c^*(i), r)) \geq 0. \end{aligned}$$

where $\Delta p_o = p'_o(p_c^*(i)', r) - p'_o(p_c^*(i), r)$, $0 < \xi \leq r'_c$. The first inequality comes from changing the integral upper limit r_c to r'_c . The second inequality is because $E_c^*(r)$ is an increasing function of r and by taking the partial derivative of $\Delta p_o \geq 0$ with respect to r_c . With tedious derivation, we can show that there exist ξ , when $r_c < \xi$, $\Delta p_o \leq 0$, resulting in $\int_0^\xi E_c^*(r) \Delta p_o dr \geq E_c^*(\xi) \int_0^\xi \Delta p_o dr$, and when $r_c \geq \xi$, $\Delta p_o \geq 0$, resulting in $\int_\xi^{r'_c} E_c^*(r) \Delta p_o dr \geq E_c^*(\xi) \int_\xi^{r'_c} \Delta p_o dr$. Thus, $\bar{E}_c(r_c)$ is an increasing function of r_c .

REFERENCES

- [1] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1801–1819, 2014.
- [2] Cisco Visual Networking, "Global mobile data traffic forecast update 2014-2019," *White Papers*, 2015.
- [3] N. Golrezaei, A. F. Molisch, A. G. Dimakis, and G. Caire, "Femto-caching and device-to-device collaboration: A new architecture for wireless video distribution," *IEEE Commun. Mag.*, vol. 51, no. 4, pp. 142–149, 2013.
- [4] N. Golrezaei, P. Mansourifard, A. Molisch, and A. Dimakis, "Base-station assisted device-to-device communications for high-throughput wireless video networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3665–3676, 2014.
- [5] M. Ji, G. Caire, and A. Molisch, "Wireless device-to-device caching networks: Basic principles and system performance," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 176–189, 2016.
- [6] J. Jiang, S. Zhang, B. Li, and B. Li, "Maximized cellular traffic offloading via device-to-device content sharing," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 82–91, 2016.
- [7] S. Andreev, A. Pyattaev, K. Johnsson, O. Galinina, and Y. Koucheryavy, "Cellular traffic offloading onto network-assisted device-to-device connections," *IEEE Commun. Mag.*, vol. 52, no. 4, pp. 20–31, 2014.
- [8] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and Zipf-like distributions: Evidence and implications," *IEEE INFOCOM*, 1999.
- [9] D. Stoyan, W. S. Kendall, J. Mecke, and L. Ruschendorf, *Stochastic geometry and its applications*. Wiley New York, 1987, vol. 2.
- [10] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [11] R. W. Heath, T. Wu, Y. H. Kwon, and A. C. Soong, "Multiuser MIMO in distributed antenna systems with out-of-cell interference," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4885–4899, 2011.
- [12] A. Papanicolaou, *Taylor approximation and the delta method*, 2009. [Online]. Available: <http://web.stanford.edu/class/cme308/OldWebsite/notes/TaylorAppDeltaMethod.pdf>
- [13] H. Kim and G. de Veciana, "Leveraging dynamic spare capacity in wireless systems to conserve mobile terminals' energy," *IEEE/ACM Trans. Netw.*, vol. 18, no. 3, pp. 802–815, 2010.
- [14] Stirling formula, *Encyclopaedia of Mathematics*. [Online]. Available: http://www.encyclopediaofmath.org/index.php?title=Stirling_formula&oldid=13618